DOCUMENT RESUME

ED 281 721 SE 047 915

AUTHOR Williams, Janice E.
TITLE Recent Findings from the SIMS Study.
PUB DATE Apr 87
NOTE 15p.; Paper presented at the Annual Meeting of the Western Psychological Association (67th, Long Beach, CA, April 23-26, 1987).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Research; Grade 8; *International Studies; Junior High Schools; *Mathematics Achievement; *Mathematics Instruction; Secondary Education; *Secondary School Mathematics; *Time Factors (Learning)
IDENTIFIERS *Mathematics Education Research; *Second International Mathematics Study

ABSTRACT

This paper presents an in-depth study of the effect of opportunity to learn (OTL) on eighth-grade students based on mathematics performance in an international setting. Item screening reduced the original 199 items from the Second International Mathematics Study (SIMS) to 152 items for which both OTL and p-values were reported across all items for eight participating countries. Initial analyses indicated no overall country significant relationship between OTL and student achievement. In addition, follow-up analyses of the Country by Item residuals revealed only minor overall relationships. This conclusion was substantiated with both correlational and ANOVA methodology. It was concluded that these data provided no real support for hypotheses relating OTL to differences in student performance between countries. Two alternative methods of measuring OTL were suggested. (Author/MNS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Janice E. Williams
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED281721

RECENT FINDINGS FROM THE SIMS STUDY

Janice E. Williams

University of California, Los Angeles

A paper presented to the Western Psychological Association, Long Beach, California on April 23, 1987

An international concern for mathematics education has led to two major cross-national research projects. Results from both the International Project on the Evaluation of Educational Achievement [IEA] (Husen, 1967) and the Second International Math Study [SIMS] (Wolfe, 1979) have been used to compare student mathematics achievement across countries.

One major finding from the IEA study indicated that students had higher performance levels in countries where the teachers considered the tests to be more appropriate to the experiences of their students. Building upon this finding, results from the more current SIMS revealed that some of the variation among student achievement scores could be attributed to the differences between students' "opportunity to learn" the material that was tested. Conclusions from cross-national comparisons are strengthened when assurances can be made that test items are derived from topics the students are familiar with (Stigler, Lee & Stevenson, 1982; Stevenson et al., 1985).

Opportunity to learn (OTL) has been called time-on-task, academic learning time and involved time in the literature

SE047915

BEST COPY AVAILABLE

(Borg, 1980; Caldwell, Huitt & Graeber, 1982; Frederick & Walberg, 1980; Karweit, 1980). OTL has been conceptualized as a predictor of student achievement, a mediator between teacher behavior and student achievement, and a behavioral indicator of on-going student learning (Cornbleth, 1980). OTL also has a potential for influencing efforts to enhance instructional effectiveness. Even on an intuitive basis, most people appear to agree that OTL is an important determinant of academic achievement.

It would appear then, that an appropriate analysis of SIMS data would include OTL as a major variable in assessing student achievement. The current study explored the effect of teacher reported OTL and student mathematics achievement across eight countries.

METHOD

OTL appears to define the difficulty level of a particular country's test, relative to other countries. In addition, the level of difficulty appears to depend on the degree of item overlap between tests defined by different countries' OTL. A higher performance could be expected when there is a high degree of overlap between the items on one country's test defined on its own OTL to tests in other countries. Thus, a portion of performance attributed to differences between countries may be due to item overlap among OTL country defined tests (L. Burstein, personal communication, 1986).

It seems clear that this overlap issue may affect the conclusions from cross-national survey research. Therefore, only

a specific portion of the original SIMS questionnaire item pool were retained for the current analyses. Of the 199 original items, 19 of these were used only by Japan and were deleted. With 180 items remaining, 23 were used by every country but Japan. These were also excluded, leaving 157 test items. British Columbia had five items for which they reported p-values but no OTL. These five items were also deleted from the analyses. Thus, all of the items (N=152) where every country reported full p-value and OTL data were selected for inclusion. This procedure allowed for analyses based upon all items, rather than just those items meeting a specific country's predetermined OTL criterion level. In order to avoid the overlap issue, all subsequent analyses were conducted on every item, country by country.

The current study focused on the 8th grade sample from the full, 8 country (Belgium Flemish, Canada British Columbia, Canada Ontario, France, Japan, New Zealand, Thailand and the United States), longitudinal SIMS study. Data came from two primary sources. Mathematics achievement test scores for students in each country were used in calculating p-values, or the proportions of students passing each of the 152 items (Wolfe, 1983; Chang, 1985). All test items were in multiple choice format with five answer options per item. These items were distributed across two instruments, a core test and a set of four rotated forms. While all students completed the core items, the rotated forms were randomly distributed within classrooms. Five mathematics subscore areas were tested; arithmetic, algebra, geometry, statistics and measurement.

The OTL data were computed from the IMLD20 tape, version 2.0 of the IEA Second Math Longitudinal Databank (Wolfe, 1986). Only one OTL score per test item was used. OTL was computed by collapsing two items from the teacher surveys. Teachers in each country indicated whether or not they would teach or review the mathematics needed to correctly answer each item in their classroom during the coming school year. This information was combined with the same teacher's response to whether or not students had previously been taught the material needed to answer each item correctly. Thus the OTL data, as reported by the teachers, identified those items that had been previously taught, would be taught in the coming school year, or both, in each country.

RESULTS

Initial analyses of variance of both OTL and student performance p-values (PVAL) were conducted with items (152) completely crossed with countries (8). These analyses are shown in Table 1. As noted there, although significant country to

---

Insert Table 1 about here

---

country differences did exist for OTL, these were relatively small (a range of 0.61 to 0.78). The computed value of ω^2 was only 0.03, which is considered to be a small effect (Cohen, 1977).

The significant country to country differences for PVAL have been well documented in the literature. The data in this analysis indicated a similar hierarchical achievement structuring. Japan, Belgium Flemish and France had relatively high overall levels of student performance compared to New

Zealand, the United States and Thailand.

A non-significant (r=0.22) correlation existed between a country's overall OTL and their overall (PVAL) performance level. Japan had the highest mean values for both, while Belgium ranked second in student performance but last in overall OTL. Thailand, however, had the second highest OTL but the lowest p-values. These data indicated that overall OTL was unrelated to overall performance. Such a finding is quite unexpected if OTL has a major influence on student performance.

The possibility still remained that differences in OTL, item by item, might well be responsible for major differences in performance within a particular country. As a consequence, residuals from the PVAL analysis (the Country x Item source) were explored to determine whether or not differences in OTL accounted for the differences in country by item performance. These residuals were adjusted for overall item difficulty. The overall PVAL was computed for each item across all countries. Each item was then equated in terms of overall difficulty by subtracting out that portion of PVAL common to all countries.

The 1216 residuals were approximately normally distributed about zero with a skewness of only -0.2. This indicated that transformations of the data (arc-sine, logit, p-to-z, etc.) were unnecessary. These residuals revealed the discrepancy between a particular country's student level of achievement on a particular item compared to the expected performance, given the overall item difficulty level and that country's overall performance level. It is reasonable to assume that the unique portion of these

differences should be due to variations in the level of OTL. If a country does not teach a particular item (a low OTL value), then students from that country would be expected to do fairly poorly on that item (a negative residual). Similarly, high OTL values should be associated with positive residuals.

The correlation between all residuals and the corresponding OTL values was statistically significant (r=0.19). However, this is of little practical significance, since less than 4% of the variation in the residuals could be accounted for by variations in OTL. Such a finding might be partially due to the small but significant country to country differences in overall OTL level. It would appear that a reasonable approach would be to analyze these data country by counry, rather than as a group. To assess this possibility, separate correlations between the 152 residuals and corresponding OTL values were computed for each country. These results, shown in Table 2, were quite unexpected. It

---

Insert Table 2 about here

---

appears that the "OTL effect" was country-specific. One possible explanation of these results was that a "critical" value of OTL existed. That is, a minimum number of students must have been exposed to the material before any higher level of performance on an item was seen. If so, then treating OTL as a continuous variable (as in Table 2) is an inappropriate method for analyzing these data.

To assess this possibility, mean p-values were computed for each of 10 different levels of OTL (.00-.09, .10-.19, etc.). A plot of these mean p-values by each level of OTL is shown in

Figure 1. As noted there, some support for the "critical value

---

Insert Figure 1 about here

---

hypothesis" existed. The function did not appear to be monotonic. Throughout the vast mid-range of OTL values (from 0.20 to 0.69) changes in OTL appeared to have little or no effect upon student performance. A slight increase in performance was apparent at the 0.7 OTL value. Performance increased again at 0.8 and then markedly increased at the 0.9 level. It should be noted that a plot of sample size at each OTL level produced a steadily increasing function.

Consequently, overall means for each country adjusted for item difficulty were plotted for both OTL "Yes" (at or exceeding the OTL criterion) and OTL "No" (below the OTL criterion) for each of the potential critical values (0.7, 0.8 and 0.9). In each computation only the OTL value for that country was used as the criterion to avoid potential overlap problems. Again it appeared that a positive OTL "Yes" value for an item should be reflected in a positive residual while a negative OTL "No" value should be associated with a negative residual.

At the .01 level of significance (selected to correct for multiple comparison errors) unequal-n ANOVA's on the points from all three plots supported the major findings of the previous correlational analysis (Table 2). Two countries (Belgium and France) showed significant differences between OTL "Yes" and "No" on the item adjusted residuals. In addition, Japan, New Zealand, Thailand and the United States showed no significant differences

between the two OTL categories. For the Canadian countries (British Columbia and Ontario), whether or not significant differences were obtained depended upon the OTL criterion level. When evaluated at the 0.7 critical value, only British Columbia exhibited significant "Yes" vs. "No" differences. At this critical level of OTL, the data reflect exactly the same pattern of results presented in Table 2. It appears that the most reasonable conclusion from these data is that the OTL effect was country-specific.

DISCUSSION

Although the hierarchical structuring of student mathematics achievement by country (i.e., Japan on top) was supported by these data, it appears that previous OTL findings were not replicated. Specifically, in this study which controlled for item overlap, it appeared that the small country to country OTL value differences were of little practical significance. Of major importance, overall OTL was unrelated to overall student performance. In addition, when relative individual item performance, by country, was considered along with OTL, only limited evidence for an OTL effect was found.

Three critical OTL cut-off values were determined to be influential, but only slight increases in the level of student performance existed at two levels, with OTL at 0.9 reflecting the highest level of student performance. This was confounded, however, by the high OTL ratings given most items by the majority of teachers and the relatively even spread of p-values across this 0.9 OTL value.

The correlation of residuals and OTL, for each country, and the plots of overall item-adjusted means for each country, for both OTL "Yes" and "No" indicated that the OTL effect was, in fact, country specific for these data. The current study, then, showed no real effect of OTL as it was currently measured on differences in student performance in the eight countries.

Perhaps "OTL" could be better measured by having teachers also estimate the proportion of their students who will give the correct answer to an item. In such a measurement, the variable of "exposure" could be differentiated from "learning." Alternatively, it might be possible to randomly sample a cohort of students who would not actually take the test but would only relate their perception of whether or not OTL occurred for their class. Either alternative may be preferable to the current method of measurement.

References

Caldwell, J. H., Huitt, W. G., & Graeber, A. O. (1982). Time spent in learning: Implications from research. The Elementary School Journal, 82, 471-480.

Chang, L. (1985). Second International Mathematics Study. (Tech. Rep. No. I). Illinois: Stipes Publ. Co.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

Cornbleth, C. (1980). Questioning time on task. Viewpoints in Teachning and Learning, 56, 107-113.

Crosswhite, J. F., et al. (1985). United States Summary Report. Second International Mathematics Study. Illinois: Stipes Publ. Co.

Frederick, W. C., & Walberg, H. J. (1980). Learning as a function of time. Journal of Educational Research, 73, 183-194.

Gettinger, M. (1984). Achievement as a function of time spent in learning and time needed for learning. American Educational Research Journal, 21, 617-628.

Husen, T. (1967). International Study of Achievement in Mathematics: A Comparison of Twelve Countries. New York: Wiley.

Karweit, N. L. (1980). Time in school. In Research in Sociology of Education and Socialization: An Annual Compilation of Research. Greenwich, CT: JAI Press.

Stigler, J. W., Lee, S., Lucker, G. W., & Stevenson, H. W. (1982). Curriculum and achievement in mathematics: A study of elementary school children in Japan, Taiwan and the United States. Journal of Educational Psychology, 74, 315-322.

Wolfe, R. (1983). Second Study of Mathematics. (Tech. Rep. No. I). Processing of selected longitudinal data files. Urbana, IL: International Assn. for the Evaluation of Educational Achievement.

Wolfe, R. (1986). IEA Second Math Longitudinal Databank [Machine-readable data file]. Ontario: Institute for Studies in Education (Producer). Toronto: Computer Services Group (Distributor).

Table 1

OTL and PVAL ANOVA's

		OTL		PVAL	
Source	df	M.S.	F	M.S.	F
Country	7	0.42	11.49*	0.62	66.61*
Item	151	0.32	8.74*	0.23	25.30*
Country x Item	1057	0.04	----	0.01	-----

* $p < .01$

Means:	BFL	BC	ONT	FRA	JPN	NZ	THA	USA
OTL	0.61	0.70	0.71	0.72	0.78	0.64	0.73	0.68
PVAL	0.57	0.53	0.49	0.53	0.62	0.46	0.42	0.47

Table 2

Correlations of Country Residual by OTL

Country	r
Belgium Flemish	0.50*
Canada British Columbia	0.29*
Canada Ontario	0.14
France	0.41*
Japan	0.09
New Zealand	-0.13
Thailand	0.10
United States	0.02

* $p < .05$

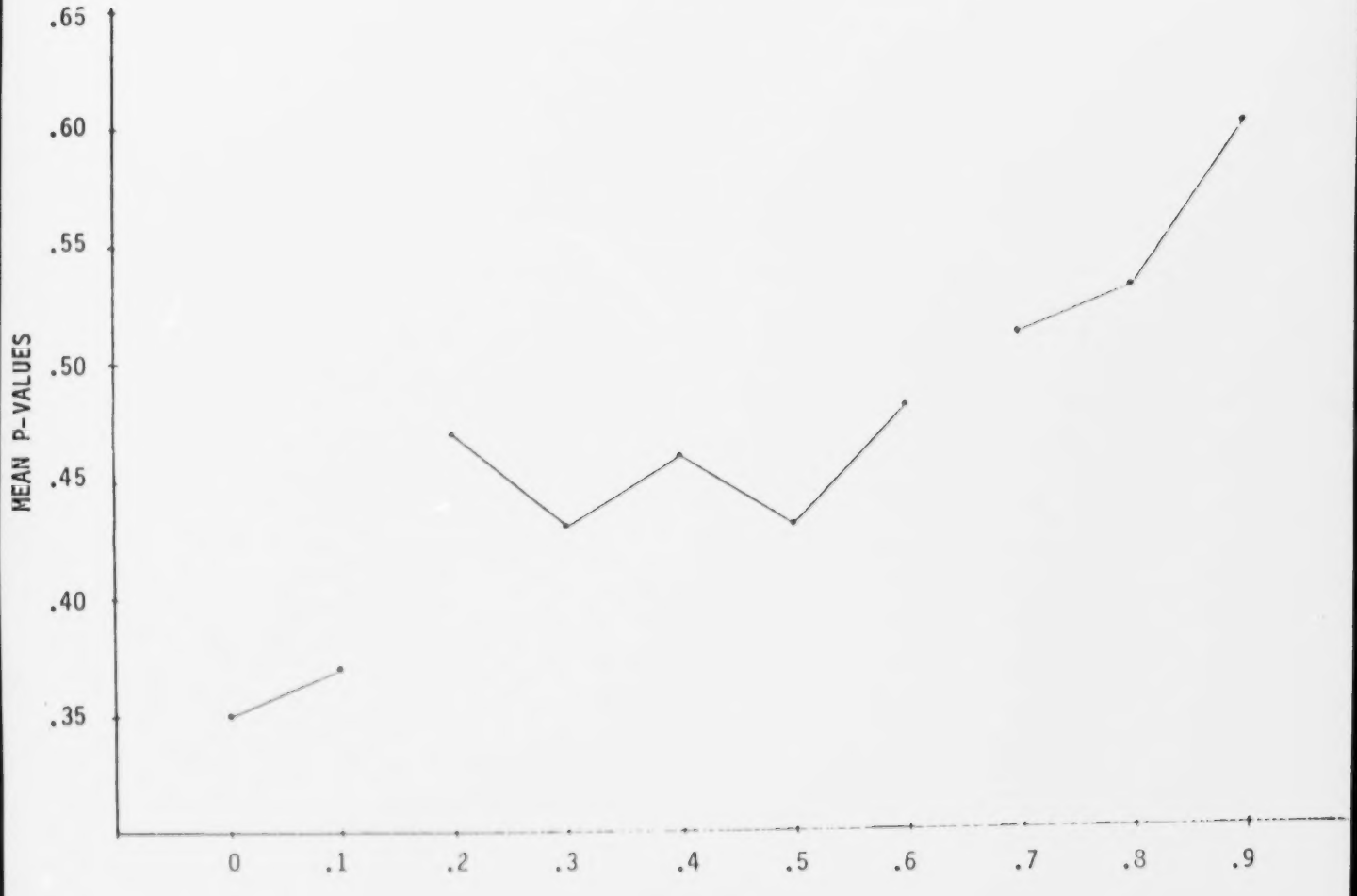

FIGURE 1